Interactions of the site-specific recombinases XerC and XerD with the recombination site dif

Nucleic Acids Research, 1994, Vol. 22, No. 25 5613–5620

Interactions of the site-specific recombinases XerC and XerD with the recombination site *dif*

Garry W.Blakely and David J.Sherratt*
Microbiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received September 20, 1994; Revised and Accepted November 18, 1994

ABSTRACT

The Xer site-specific recombination system of *Escherichia coli* is involved in the stable inheritance of circular replicons. Multimeric replicons, produced by homologous recombination, are converted to monomers by the action of two related recombinases XerC and XerD. Site-specific recombination at a locus, *dif*, within the chromosomal replication terminus region is thought to convert dimeric chromosomes to monomers, which can then be segregated prior to cell division. The recombinases XerC and XerD bind cooperatively to *dif*, where they catalyse recombination. Chemical modification of specific bases and the phosphate–sugar backbone within *dif* was used to investigate the requirements for binding of the recombinases. Site-directed mutagenesis was then used to alter bases implicated in recombinase binding. Characterization of these mutants by *in vitro* recombinase binding and *in vivo* recombination, has demonstrated that the cooperative interactions between XerC and XerD can partially overcome DNA alterations that should interfere with specific recombinase–*dif* interactions.

INTRODUCTION

Prokaryotic cell division requires that newly replicated monomeric chromosomes are segregated to distinct cellular locations before cell division can be completed (1). Odd numbers of homologous recombination events between replicating, or newly replicated, circular chromosomes can lead to production of chromosome dimers, which must be resolved to monomers prior to nucleoid segregation. The *dif* site, present in the *Escherichia coli* K12 DNA replication terminus region, has been implicated in normal chromosome segregation through its role as a site-specific recombination locus. It has been proposed that recombination between two *dif* sites in a chromosome dimer converts the dimer to monomers. Although *dif* is dispensable, its absence produces a sub-population of filamentous cells that contain aberrant nucleoids (2, 3). We propose that the odd number of homologous exchanges that generate dimers occurs rarely, as do homologous exchanges that convert dimers to monomers. The same phenotype is demonstrated by cells that contain mutations in *xerC* or *xerD*, two genes encoding site-specific recombinases that function in recombination at *dif*. The protein sequences of XerC and XerD share 37% identity and show them to be members of the lambda integrase family of site-specific recombinases (3, 4).

Recombination proceeds, after recombinase binding and synapsis of sites, by activation and subsequent cleavage of specific phosphodiester bonds. The active site tyrosine of the recombinase acts as a nucleophile and cleaves the DNA to form a covalent protein–DNA intermediate. Free 5′ hydroxyl ends, generated by the initial DNA cleavage, then act as attacking nucleophiles to religate the DNA. A total of four strand cleavages and religations are required to generate recombinant products (reviewed in 5).

The XerC recombinase was initially identified by its role in resolution of ColE1 plasmid multimers (generated by homologous recombination) to monomers. This recombination is necessary for the stable inheritance of this naturally occurring high copy number plasmid and its relatives (6, 7). A second recombinase, XerD, was identified by sequence homology to XerC and is encoded in an operon with *recJ* and *dsbC* (8–10). XerC/XerD-mediated site-specific recombination *in vivo* at the 210 bp *cer* locus of ColE1 shows selectivity for intramolecular resolution, i.e. recombination only occurs between two sites in direct repeat in the same molecule (usually a dimeric plasmid) to produce two monomers. *cer* consists of a 30 bp core sequence to which XerC and XerD bind (Figure 1) and ~180 bp of upstream accessory sequences (11). Recombination at *cer* also requires two accessory proteins, ArgR and PepA. The precise function of these proteins is not known, but they and the *cer* accessory sequences have been implicated in resolution selectivity (12). *cer*-like sequences present in related, naturally occurring plasmids are also required for plasmid stability (4).

A DNA fragment of 32 bp which contains a functional *dif* site is sufficient to allow Xer-mediated plasmid multimerization and dimer resolution (3). The structural organization of *dif* is similar to that of related lambda integrase family site-specific recombination loci, e.g. P1 *loxP*, and *Saccharomyces cerevisiae* FRT (13, 14). The core recombination site is divided into two 11 bp half-sites, which flank a 6 bp central region. The half-

*To whom correspondence should be addressed

sites are related by partial dyad symmetry and show homology with the core sequence of *cer* (Figure 1). Copper–phenanthroline footprinting demonstrated that XerC binds to the left half-site while XerD binds to the right half-site; binding of both proteins is highly cooperative. A requirement for two different recombinases appears to provide the asymmetry for ensuring correct alignment of recombining sites before the first strand exchanges occur. Neither half-site can be used to replace the other half-site and the putative active sites of both proteins are required for recombination *in vivo*, demonstrating the requirement and involvement of both proteins in the strand cleavage and transfer reactions (8). The novel requirement for two recombinases may enable each pair of strand exchanges to be under separate genetic control, as well as facilitating site alignment immediately after site replication in the chromosome. Based on the dyad symmetry of the half-sites and by analogy with the cleavage positions from other recombinases (13, 14), the boundaries of the central region and recombinase binding sites have been proposed to contain the bases involved in strand nicking and exchange (15). In *dif* the central region contains 6 bp, while in *cer* it consists of 8 bp; this difference may constitute a major determinant of the outcome in the recombination reaction and the requirement for accessory sequences.

The presence of limited dyad symmetry as part of the 'inner' sequence of each binding site suggests that the 'outer' sequence specifies the recognition differences between the XerC and XerD binding sites. Because of the dyad symmetry and the requirement for two related recombinases, a study of the DNA sequence and structural requirements which determine specificity for interaction between XerC, XerD and *dif* was undertaken. Footprinting techniques using the DNA modifying reagents dimethyl sulphate, potassium permanganate and *N*-ethyl-*N*-nitrosourea provide evidence for protein interactions with guanine and thymine bases in the major grooves, adenines in the minor grooves and phosphates in the backbone of DNA within the *dif* site. Selected base pairs identified within the sequence were mutagenized to determine the effect of base changes on recombinase binding and *in vivo* recombination.

MATERIALS AND METHODS

Plasmids, DNA fragments and Xer proteins

DNA fragments containing the functional *dif* site were generated from plasmid pMIN33 (2). The 67 bp *Hin*dIII–*Kpn*I fragment was used to end-label the 3′ recessed end of the top strand using the Klenow fragment of DNA polymerase I; the 73 bp *Eco*RI–*Sph*I fragment was similarly used to end-label the bottom strand. XerC and XerD proteins were partially purified by fast protein liquid chromatography on Mono S columns as previously described (8).

Complementary synthetic oligonucleotides, corresponding to *dif* sites with point mutations at positions within either the XerC binding site (C–8 or A–7) or the XerD binding site (T4 or G8; see Figure 1), were annealed and ligated into the *Xba*I and *Sal*I sites of pUC18 to give plasmids pGB210, pGB211, pGB212 and pGB213 respectively. Sequences of cloned mutant *dif* sites were confirmed by the dideoxy chain termination method. Reporter plasmids, containing direct repeats of each mutant *dif* site flanking a kanamycin resistance gene, were made as previously described for construction of the wild-type *dif* reporter plasmid pSDC124 (3). DS941 (Xer^+) and DS984 ($XerC^-$) are *recF* derivatives of AB1157 (3). Genetic testing for loss of reporter plasmid kanamycin resistance was performed after isolation of plasmid DNA from DS941. Subsequent transformation of recombination products into DS984 allowed screening for sensitivity to kanamycin; 100 transformants were tested for each reporter plasmid.

DMS methylation interference

Labelled DNA and 1 μl of poly(dI–dC) (1 mg/ml) were dissolved in 200 μl of 50 mM sodium cacodylate, pH 8.0, and 1 mM EDTA. 1 μl of DMS was added and the DNA modification reaction proceeded for 5 min at room temperature before termination by spun column chromatography through Sephadex G50 followed by precipitation. XerC and XerD extracts were then added to the modified DNA to give a final protein concentration of 3.4 μg/ml, under binding conditions previously described (8). Electrophoresis through 6% polyacrylamide gels was used to separate bound from unbound DNA, with protein–DNA complexes detected by autoradiography. After elution from gel slices, the labelled DNA was cleaved at modified positions using 1 M piperidine at 90°C for 30 min, lyophilized, resuspended in loading dye and analysed by electrophoresis through 20% sequencing gels.

Potassium permanganate interference

Labelled DNA was dissolved in 55 μl of 30mM Tris–HCl, pH 8, followed by denaturation at 100°C for 3 min (16). Modification was carried out by adding 60 μl of 0.25 mM $KMnO_4$ for 12.5 min at room temperature before purification by spun column chromatography and precipitation. Isolation and analysis of protein–DNA complexes and free DNA was as described above. Strand cleavage was performed in 1 M piperidine at 90°C for 30 min.

Ethylation interference

Labelled DNA was dissolved in 100 μl of 50 mM sodium cacodylate, pH 7.0, 1 mM EDTA and 1 μg of poly(dI–dC). 100 μl of a saturated solution of *N*-ethyl-*N*-nitrosourea in ethanol was added and incubated at 50°C for 1 h. After the modification, the DNA was precipitated with 0.3 M sodium acetate and 2 volumes of ethanol. Isolation and analysis of protein–DNA complexes and free DNA was as described above. Strand cleavage was performed in 100 mM sodium hydroxide at 90°C for 30 min (17). Samples were mixed with urea loading dye and loaded directly onto 20% sequencing gels.

Quantitation

β-Particle emission from ^{32}P-labelled DNA fragments in sequencing gels was measured using a Molecular Dynamics PhosphorImager with ImageQuant software. Autoradiograms were scanned using a Joyce-Loebel microdensitometer.

RESULTS

Methylation interference of XerC and XerD binding to *dif* DNA *in vitro*

Examination of core recombination sites derived from the *E.coli* chromosome and from naturally occurring plasmids shows that XerD binding sites are highly conserved, while XerC binding sites are more divergent (Figure 1). What specifies XerD recognition of the right half-sites and XerC recognition of the left half-sites at these related loci? Just three nucleotides, located in the 'outer' sequences of each half-site, distinguish the XerC

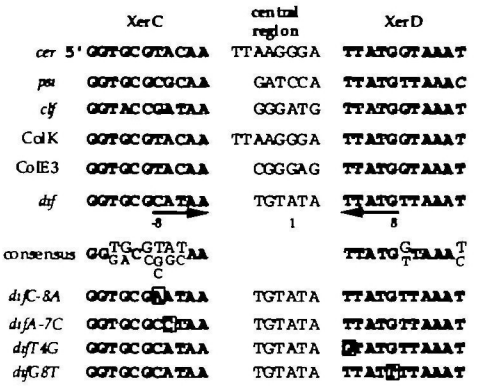

Figure 1. Alignment of homologous sequences from some known Xer recombination sites. XerC binding sites are less conserved than XerD binding sites. *cer* is from ColE1; *psi* is from pSC101; *clf* is from CloDF13; *dif* is from the *E.coli* chromosome; the sites from plasmids ColK and ColE3 are also given (4). Base pair coordinates for *dif* are indicated. Sequences of the mutant *dif* sites used in this study are shown below the consensus. Bases that have been changed are boxed. The arrows denote the region of dyad symmetry within *dif* and are designated as the 'inner' sequences.

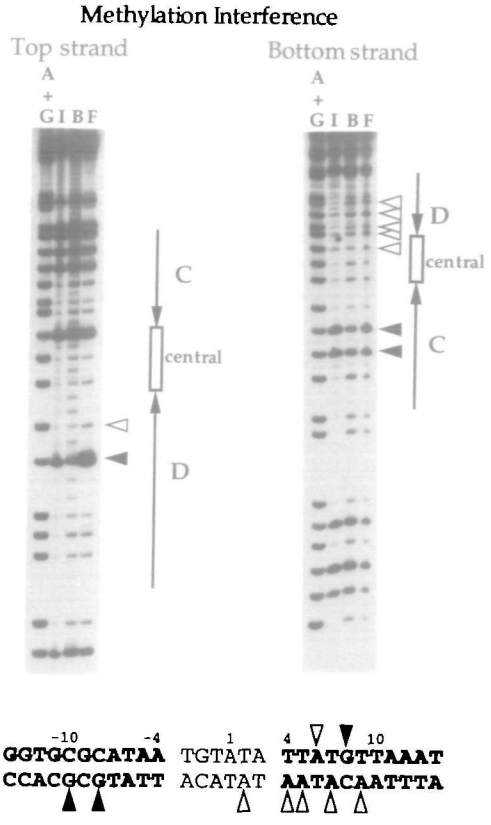

Figure 2. Methylation interference for both strands of *dif* DNA modified by DMS at N-7 of guanine and N-3 of adenine. The modified 3' labelled DNA (I) was treated with XerC and XerD and separated into bound (B) and free (F) forms. Analysis, next to a Maxam–Gilbert A + G sequencing reaction, demonstrates the sequence positions that interfere with protein binding. Interference is seen as reduction of band intensity in the bound (B) lane and an enhancement in the free (F) lane. A diagrammatic representation of *dif* is shown adjacent to corresponding portions of the sequencing gel. Positions of strong interference are denoted (▲), with weaker interference as (△).

and XerD binding sites and presumably provide the differential recognition specificity for appropriate recombinase binding. Regions of dyad symmetry between the half-sites are largely confined to the 'inner' sequences of each binding site. These may be required for specific interactions with common regions of the two proteins.

Chemical modification of DNA was used to identify regions which may act as recognition sequences. Modification interference has proved a powerful technique for identification of sites in DNA which are in close proximity to bound protein in specific complexes (18). Dimethyl sulphate (DMS) methylates the N-7 of guanine and the N-3 of adenine in the major and minor grooves of double stranded B-form DNA respectively (17). The N-7 of guanine and the N-3 of adenine can donate hydrogen bonds to amino acids (19); methylation of these positions can interfere with protein–DNA complex formation where specific hydrogen bonding is required. Addition of protein to modified DNA followed by separation of bound and unbound forms allows determination of bases that interfere with binding using a methodology derived from Maxam–Gilbert DNA sequencing (20).

DNA fragments containing the *dif* sequence labelled at the 3' end were modified with DMS and then incubated with XerC and XerD. Bound and unbound DNA were separated using non-denaturing polyacrylamide gel electrophoresis followed by analysis on denaturing polyacrylamide gels after DNA cleavage. Modified bases which interfere with binding are under-represented in the bound fraction of DNA and are enriched in the free fraction of DNA. These interference experiments were only informative when used with XerC and XerD combined, because of the low affinity of the individual proteins for the binding sites. Modification interference patterns were only observed when subsaturating amounts of XerC and XerD were present, suggesting that modification did not abolish binding, presumably because of the cooperative nature of XerC and XerD interactions. In general, methylation of G residues interfered with recombinase binding more strongly than A methylation. Note that piperidine-mediated cleavage of methylated guanine is more efficient than cleavage of methylated adenine (reviewed in 17).

Within the 31 bp of the *dif* sequence protected by XerC and XerD from attack by 1,10-phenanthroline–copper (8) there are 6 guanine residues in the top strand and 2 guanine residues in the bottom strand (Figure 1). Of the 6 G residues in the top strand, 4 are in the XerC binding site, while there is one each in the XerD binding site and the central region. The only methylated guanine in the *dif* top strand that interfered with formation of the XerC/XerD complex was at position 8 in the XerD binding site (Figure 2). Analysis of the bottom strand indicated that two modified G residues interfered with XerC/XerD complex formation; these were at positions −8 and −10 in the XerC binding site (Figure 2).

Minor interference from a modified adenine was detected on the top strand at position 6. Adenines on the bottom strand which appeared to be involved in binding were observed at positions 2, 4, 5, 7 and 9; these were all detected as minor enhancements in the unbound DNA fraction. The G residues at positions 8 and −8, top and bottom strands respectively, are in equivalent positions within the limited dyad symmetry of the *dif* sequence (Figure 1). Most of the interference positions map in the XerD

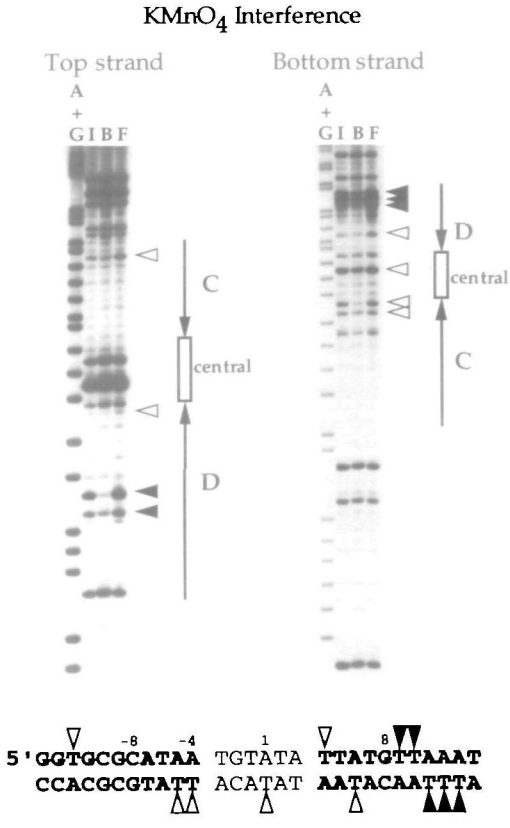

Figure 3. Modification of *dif* DNA using $KMnO_4$ was used to identify thymine major groove contact points important for XerC and XerD binding. Modified input DNA (I), bound complex (B) and free DNA (F) are shown, after cleavage with piperidine, adjacent to a Maxam–Gilbert A + G sequencing reaction. Both DNA strands were analysed and are shown with the corresponding regions of *dif* next to the appropriate sequence. Positions of strong interference are denoted (▲), with weaker interference as (△). Note the hypermodification of T2 in the central region and hypomodification of T−6, T4 and T5 in the dyad symmetry of the top strand.

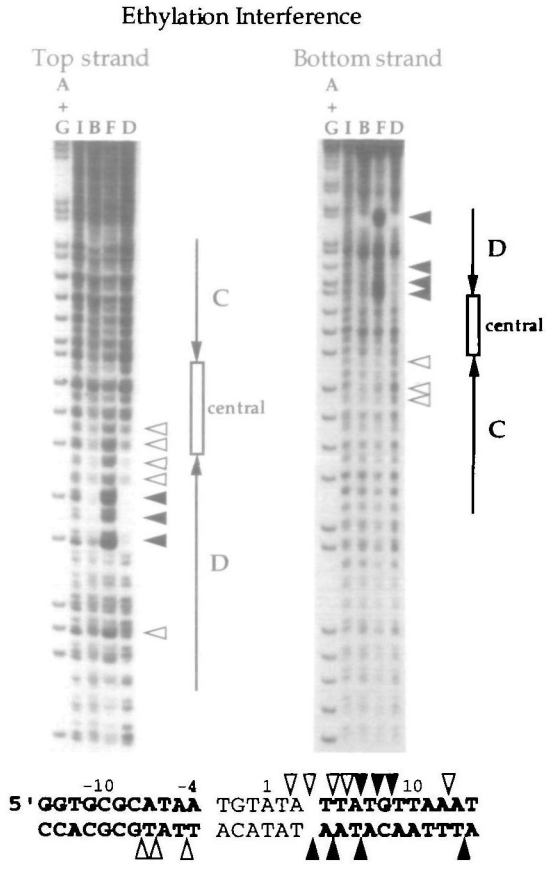

Figure 4. Ethylation of phosphates in the DNA backbone of *dif* leads to interference with XerC and XerD binding. Separation of input DNA (I) from XerC/XerD bound complex (B) and free DNA (F) allows detection after alkaline cleavage A sample derived from a presumptive XerD bound complex (D) is also shown The D lane shows strong interference in the right half-site. The only apparent interference positions in the left half-site were seen as enhancements at base pairs −4, −6 and −7 in the bottom strand (lane D). Lanes marked A + G are adenine and guanine Maxam–Gilbert sequencing reactions. Note the doublets produced from cleavage 3′ and 5′ of the modified phosphates. The strong (▲) and weak (△) interference positions are indicated on the sequencing gel and the corresponding DNA sequence.

binding site; the higher affinity of XerD compared to XerC may account for this. In no case did modification of any A residues or G residues completely block protein–DNA complex formation; indeed, in the presence of saturating amounts of XerC and XerD all modified DNA electrophoresed as bound complex (data not shown).

Interference of XerC and XerD binding to *dif* DNA modified with potassium permanganate

The non-polar methyl group of thymine at position C-5 is important for discrimination between thymine and cytosine and provides a surface for van der Waal's contacts with amino acids (19, 21). The 5,6 double bond of thymine residues in single-stranded DNA can be selectively oxidized by potassium permanganate to produce the glycol form. Thymine glycols have been shown to prevent protein binding at specific DNA sequences, for example steroid hormone receptor interactions with promoter regions and FLP recombinase binding of FRT sites (16, 22).

End-labelled *dif* DNA was denatured and treated with $KMnO_4$ at a concentration appropriate for glycolization. After reannealing, the modified DNA was incubated with subsaturating amounts of XerC and XerD before separation on non-denaturing polyacrylamide gels. Bound and free DNA were analysed on sequencing gels after cleavage with piperidine.

Analysis showed that modification of thymines within the *dif* sequence, in both top and bottom strands, interfered with binding of XerC and XerD. The top strand of *dif* showed four positions of interference when thymines were modified at positions −12, 4, 9 and 10 (Figure 3). Not all thymines in the sequence were oxidized equally. Bases between the regions of dyad symmetry were always undermodified in comparison to sequences which constitute the rest of the binding site. The thymine at position 2 in the central region was hypermodified. The reproducibility of this modification pattern suggests it is a function of the DNA sequence and may be a consequence of secondary structure, e.g. a hairpin loop occurring during the modification procedure. Similar anomalies have been reported for $KMnO_4$ modification of certain thymine residues in FRT DNA, where the protein binding sites are almost perfectly symmetrical. Secondary

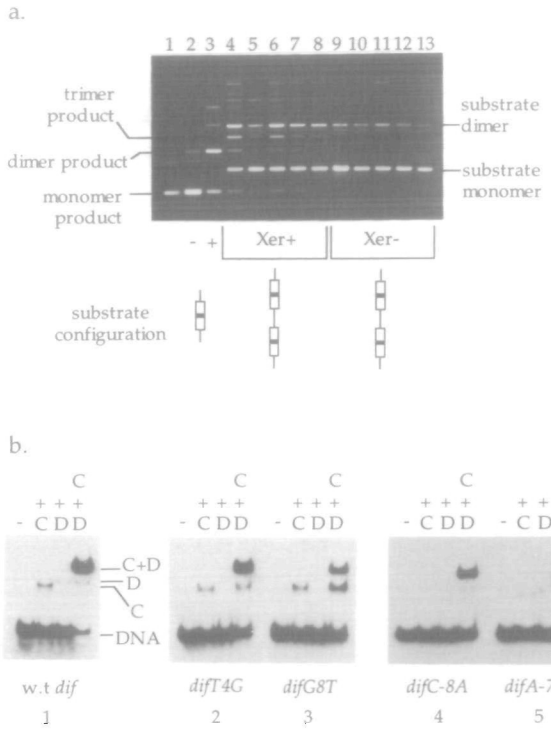

Figure 5. (a) Intra- and intermolecular recombination *in vivo* between *dif* sites. Lane 1, pUC18 monomer marker; lane 2, pMIN33 from Xer$^-$ strain; lane 3; pMIN33 from Xer$^+$ strain; lane 4, *dif* reporter from Xer$^+$ strain; lane 5, *difC−8A* reporter from Xer$^+$ strain; lane 6, *difA−7C* reporter from Xer$^+$ strain; lane 7, *difT4G* reporter from Xer$^+$ strain; lane 8, *difG8T* reporter from Xer$^+$ strain; lane 9, *dif* reporter from Xer$^-$ strain; lane 10, *difC−8A* reporter from Xer$^-$ strain; lane 11, *difA−7C* reporter from Xer$^-$ strain; lane 12, *difT4G* reporter from Xer$^-$ strain; lane 13, *difG8T* reporter from Xer$^-$ strain (b). Gel retardation assay to determine effect of single base changes on binding of XerC and XerD to mutant *dif* sites. Comparable amounts of radiolabelled DNA were treated with equal concentrations of XerC, XerD or XerC + XerD and complexes separated by polyacrylamide electrophoresis. Note the highly cooperative binding of XerC and XerD to wild-type *dif* (panel 1). Binding affinity appeared to be reduced for mutants T4G, G8T and C−8A (panels 2, 3 and 4); compare the amounts of retarded C + D complex to free DNA.

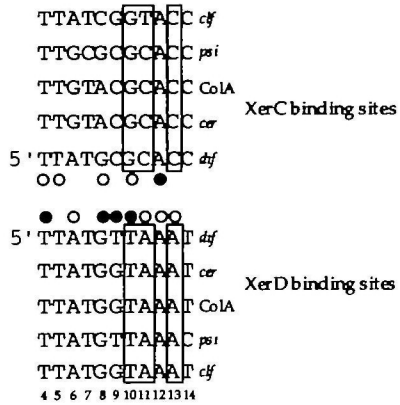

Figure 6. Alignment of half-sites from some of the known recombination loci (4) demonstrates that only three base pair positions (boxed) are unique to each half-site. This is supported by comparison of the consensus sequences for each half-site (see Figure 1). These base pairs may provide the specificity for recognition and binding of XerC and XerD and may be key determinants which distinguish between the half-sites. The bases that interfere with protein binding when modified are indicated for top (●) and bottom (○) strands. Three of the seven interference positions determined for the XerD binding site map within the unique sequence. Central regions occur 5′ to the shown sequences. Base pair coordinates are given below the last sequence.

structure in single-stranded DNA may lead to ambiguous assignment of binding interference for particular bases (22). An aberrant cleavage product which ran between bases A3 and T4 may have been a result of the hypermodification of bases between the regions of dyad symmetry.

Modification of the *dif* bottom strand produced seven thymine residues that interfered with binding; T at positions −5, −4, 1, 6, 11, 12 and 13 (Figure 3). Hypermodification of T1 in the central region of the bottom strand was also observed. Thymines at positions 9 and 10 in the top strand and 11, 12 and 13 in the bottom strand could therefore define the sequence which XerD recognizes as an appropriate binding site.

Interference of XerC and XerD binding by phosphate ethylation of the DNA backbone

Hydrogen bonding to DNA phosphodiesters can be important for positioning of specific protein structures, such as alignment of α-helices to allow ionic interactions with appropriate base pairs in adjacent major grooves (19). DNA conformation is also sequence-dependent, so proteins could indirectly recognize a specific DNA site by the arrangement of the phosphate−sugar backbone (reviewed in 23). Interactions between proteins and the DNA backbone can be studied by modifying phosphates with the addition of an ethyl group donated from *N*-ethyl-*N*-nitrosourea. The ethyl group may sterically interfere with the protein or may inhibit binding by altering the charge on the DNA backbone (24).

The alkaline cleavage chemistry for ethylated DNA produces two products from a single modification; the products being dependent on whether the cleavage occurs at the 5′ or 3′ side of the phosphate (24). The resultant products migrate with an apparent difference of half a base pair on the sequencing gels used in these studies, therefore each cleavage produces a doublet.

Analysis shows that ethylation of 3 phosphates interfered strongly with binding, while ethylation of another 5 interfered weakly. The strong interference positions corresponded to 3′ phosphates at bases A6, T7 and G8 in the XerD binding site of *dif* (Figure 4). The weaker positions corresponded to phosphates 3′ to bases T2, A3, T4, T5 and A12. The bottom strand ethylation pattern was analogous to the top strand pattern: strong interference was observed by ethylation of 4 phosphates in the right half-site, 3′ to base pairs at positions 4, 5, 7 and 14; weak interference was observed by ethylation of 3 phosphates in the left half-site, 3′ to base pairs −4, −6 and −7. The evidence for interference in the left half-site was from enhancement of cleavage products present in a weak XerD complex isolated from the gel shift. This complex showed strong interference in the right half-site, so may have been derived from a portion of modified DNA to which XerC could not bind.

Site-directed mutagenesis of *dif*

To determine if bases detected by the modification interference analysis were of significance to protein binding and site function, site-directed mutants of *dif* were constructed. Three bases

(positions C−8, T4 and G8) assigned importance as potential contact positions were changed. One base (A−7) that did not appear to interfere with protein binding when modified was changed as a control (Figure 1). Bases were changed to sequences which did not occur in other known natural recombination sites. The mutant produced for base T4, which was within the region of hypomodification by $KMnO_4$ on the top strand, would also allow confirmation of its role in protein binding. The effects of these four mutations on *in vivo* Xer-mediated plasmid recombination and *in vitro* protein binding were studied.

Reporter plasmids containing each mutant site as direct repeats, flanking a kanamycin resistance gene, were constructed and transformed into a Xer^+ strain. Recombination products were analysed by agarose gel electrophoresis after ~30 generations. Xer-mediated intramolecular recombination between *dif* sites within the reporter plasmid produces a plasmid equivalent in size to pMIN33 (Figure 5a, lane 3), while intermolecular fusion of the reporter plasmid and its resolved products produces a complex ladder of multimeric forms (Figure 5a, lanes 4−8). Sites *difC−8A*, *difT4G* and *difG8T* were all capable of inter- and intramolecular recombination (lanes 5, 7 and 8), but appeared not to be as proficient in Xer-mediated recombination as wild-type *dif*; for example, compare the amounts of monomer and dimer product in lanes 5, 7 and 8 to lane 4. In contrast, plasmid DNA containing the site *difA−7C* gave a pattern of recombinant products indistinguishable from wild-type *dif* (lane 6).

A genetic test was used to quantify the relative *in vivo* intramolecular recombination proficiency of each of the mutant *dif* sites. Total plasmid DNA from Xer^+ strains containing reporter plasmids was scored for loss of kanamycin resistance by subsequent transformation into a Xer^- strain. After ~50 generations, the reporter plasmid containing the wild-type *dif* site showed 99% resolution, while the site *difA−7C*, which appeared to have wild-type recombination characteristics, showed 76% resolution. The other mutant sites which had shown reduced resolution by gel electrophoresis produced a lower percentage of phenotypic loss; *difC−8A* 20%, *difT4G* 5% and *difG8T* 1% resolution, confirming their reduced proficiency in recombination.

To analyse the *in vitro* binding of the recombinases to each mutant site, band shift assays were carried out. A DNA fragment containing wild-type *dif* binds either XerC or XerD weakly as judged by the amount of protein−DNA complex at a given recombinase concentration (8). The same concentrations of XerC and XerD when added together converted most of the input DNA to recombinase−DNA complexes, demonstrating the highly cooperative nature of recombinase binding to *dif* (Figure 5b, panel 1). Site *difC−8A* , containing a mutation in the XerC binding site, did not produce a detectable XerC complex and showed less than 50% of the total DNA in a complex with XerC and XerD, suggestive of a reduced overall affinity of XerC and XerD for this site, presumably arising as a consequence of reduced XerC binding (Figure 5b, panel 4; note that in this gel the XerD complex is barely visible; we do not believe this is a consequence of reduced XerD binding to this mutant site since in other band shift experiments the XerD complex was evident). Sites *difT4G* and *difG8T*, which both contained mutations in the right half-site, were able to form XerC complexes but there was no evidence for the formation of stable XerD complexes (Figure 5b, panels 2 and 3). Although both of these sites could be bound by XerC and XerD, the overall apparent affinity for the DNA was reduced, presumably because of reduced XerD binding (Figure 5b, panel 2). The control site for this mutagenesis, *difA−7C*, demonstrated

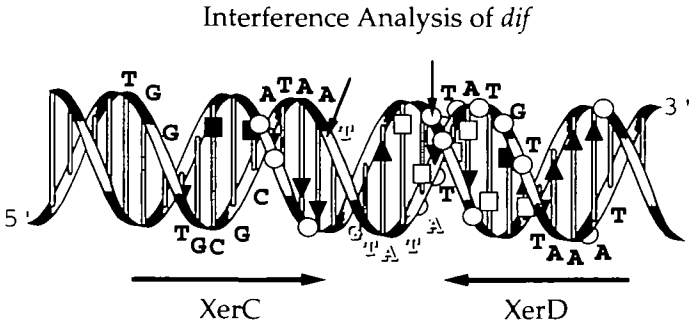

Figure 7. Representation of B-form helical DNA containing the sequence of the functional *dif* site. Base pairs are represented as vertical bars Sequence of the top strand is shown 5′−3′ adjacent to the backbone. The central region sequence is given in hollow lettering Modified guanines that interfere with protein binding are shown as ■, modified adenines as □, modified phosphates as ○ and modified thymines as ▲. The pattern of interference suggests that XerD is in contact with both faces of the helix and is interacting with base pairs in major and minor grooves in the right half-site. The small arrows indicate the proposed sites of DNA cleavage

binding patterns which were similar to the wild-type *dif* sequence (Figure 5b, panel 5).

The *in vivo* recombination results demonstrate that mutation of nucleotides implicated in protein binding, determined from interference analysis, reduced but did not abolish function of the recombination site. In parallel, XerC and XerD binding was reduced but not abolished at these mutant sites. This data also demonstrated that the assignment of position T4 as a potential contact point was correct.

DISCUSSION

This study has identified potential points of interaction between XerC/XerD and *dif* DNA and is the first to examine molecular interactions in a two recombinase system. XerC and XerD can bind to many related natural recombination sites from plasmid DNA and the *E.coli* chromosome (4). These recombination sites contain limited dyad symmetry of 4−5 bp, yet XerC only binds to the left-half site and XerD binds to the right half-site under standard *in vitro* binding assays (8). What features of the recombination locus enable these related proteins to differentiate between the half-sites?

The sequences of XerD binding sites are highly conserved, while the XerC binding sites show much greater heterogeneity. A sequence comparison of recombinase binding sites suggests that only 3 positions within the left half-site contain base pairs which do not appear in equivalent positions in the right half-site. For example, position −10 is always a G−C pair but never an A−T pair, while position 10 is always A−T (Figure 6). These base pairs must at least be part of the sequence recognition determinants for each of the recombinases. When the positions of modification interference for top and bottom strands are mapped on aligned sequences it is clear that the 3 conserved positions at base pairs 10, 11 and 13 are important for specific recognition and binding by XerD. The XerC binding site produced fewer modification interference positions, possibly because of its ability to bind DNA less specifically. One of the detected positions corresponded to base pair −10, which is part of the unique left half-site sequence. The other interference positions are primarily within the region of dyad symmetry and

in equivalent positions within each half-site, at positions T4, T−4, G8 and G−8; though note that G at position −8 is not invariant for Xer sites. Base pairs A−12 and A12 in the 'outer' sequences appear to be invariant, these equivalent positions also appear to be protein contact positions. The Xer recombinases would not be unusual in their use of a 3−4 bp recognition sequence; many DNA binding proteins use short DNA sequences for specific bond interactions between protein structures and base pairs, for example helix 3 of the POU domain from Oct-1 makes all its hydrogen bonds with the 5'-ATGC subsite (26). Binding of the 434 repressor is also specified by a 4 bp sequence at the end of the operator, 5'-ACAA, while the remainder of the 14 bp site may modulate affinity (27).

The pattern of potential base−protein interactions at *dif* is suggestive of a bipartite binding function, where each recombinase may recognize the appropriate 'outer' unique sequence from the recombination site and use the 'inner' dyad symmetry as the determinant for interaction with a common protein structure. The most similar region of XerC and XerD (73% identity) is the putative catalytic domain II, which provides the tyrosine nucleophile and is thought to be involved in phosphodiester activation and cleavage. Thymines at positions −4, −5, 4 and 6 could interact, via the major groove, with the conserved domain II. An analogous DNA structural arrangement has been proposed for the inverted repeats of IS903. The sequence recognition determinants for the transposase may be contained within the last 9 bp of 18 bp repeats and the outer sequences may be required for an additional function (28). Dissection of the related recombinase FLP, from *Saccharomyces cerevisiae*, has identified various regions of the protein required for different parts of the recombination reaction. The presence of an ~200 amino acid proteolytic cleavage product (P21) that can bind DNA but which has little effect on DNA bending and no other function (29) may support the possibility of separate protein segments interacting with different DNA domains in each half-site.

When the positions of interference for modified bases and backbone phosphates are superimposed on a projection of B-form helical DNA (Figure 7), the pattern is suggestive of XerD binding on both faces of the helix. XerD interactions with both major and minor groove residues were detected. For example, the minor groove contact at position A3 is accessible from the back of the helix, while contact at position A7 is from the front of the helix (Figure 7). Major groove contacts extend along the whole sequence of the right half-site. The proposed model for these proteins binding to *dif* would involve a 'head-to-head' disposition of protomers along the helical axis with the proteins reaching around the helix at the cleavage position. In contrast, many characterized DNA binding proteins only bind on one face of the helix, as described for interactions between RNA polymerase−T7 promoters and AraC−*araI* complexes (24, 30). The type of binding exhibited by Hin invertase may provide a more appropriate model for interpreting the data for XerC and XerD binding. Hin uses an α-helix to make contacts with a major groove and then uses adjacent β-sheets to contact minor groove residues on both faces of the helix (31). The lack of clear phosphate interference data for the left half-site of *dif* makes exact positioning of XerC difficult. The absence of strong ethylation interference may suggest that XerC binding (in the presence of XerD) does not require specific phosphate contacts with the DNA backbone. Binding of XerC to *dif* is relatively weak compared to binding of XerD, while binding of XerC to the left half-site of *cer* is difficult to observe. XerC is also able to bind a plasmid DNA sequence adjacent to an isolated right half-site when XerD is bound (8). A combination of low specificity and highly cooperative binding may prevent steric hinderance by ethylated phosphates, which will reduce sensitivity for detection of interference positions.

The methylation and oxidation interference data, however, suggest that XerC is contacting bases in major grooves from both faces of the helix. Bottom strand contacts to T−4 and T−5 would be from the back of the helix, while contacts to G−8 and G−10 would be from the front of the helix (Figure 7). This concept of XerD and XerC binding is consistent with the proposed model for binding of the FLP recombinase, which suggests that FLP contacts both faces of the helix (22). Ethylation interference of FLP binding shows a similar array of contact positions around the site of DNA cleavage (14). Ethylation interference experiments identify potential phosphate contacts on the 'outside' of the DNA (24). The presence of contacts on one helical side has been proposed as important for allowing positioning of the protein to enable access to adjacent major grooves (32). For *dif*, contacts on both faces of the helix would explain how XerD is able to interact with base T4 through the major groove. Such a wrapping of the DNA by the protein may represent an anchoring position for maintaining DNA contact after strand cleavage. Two of the phosphate interference positions present in the right half-site of *dif* occur at the region for bottom strand cleavage, between bases 3 and 4. It has been proposed previously that one of the conserved arginine residues from the integrase family members could provide ionic interactions between the recombinase and phosphates in the DNA backbone (33). The Arg−His−Arg triad has been implicated in the activation of appropriate phosphodiester bonds prior to strand cleavage (34). The interference positions identified around the cleavage position may be the phosphates which are contacted by the catalytic domain of XerD, a function which is required for phosphodiester activation. The bases that interfere with binding when modified are situated in major grooves flanking positions of backbone phosphate contacts. No obvious DNA binding motifs are evident from analysis of the amino acid sequences of XerC, XerD or any other member of the integrase family (35).

The need to use both recombinases in the interference assays, along with their highly cooperative interactions, may have reduced the sensitivity of our analysis. Nevertheless, the site-directed mutagenesis of *dif* clearly demonstrates that the interpretation of the interference analysis was correct. Although all the mutant sites were capable of *in vivo* recombination, the amount of resolution as observed by genetic testing appears to relate to the extent of interference. Mutant *difG8T* produced less resolution in comparison to the other mutant sites; this base also produced the strongest reproducible interference signal of any base in the *dif* sequence. It therefore appears that the cooperative binding of XerC and XerD can partially overcome the reduced binding affinity for a particular mutant half-site, allowing the formation of a protein−DNA complex that can give rise to a productive synapse. Reduced stability of such a complex may explain the reduced recombination rates observed.

REFERENCES

1. Rothfield, L. I. (1994) Cell 77, 963−966.
2. Kuempel, P. L., Henson, J. M., Dircks, L., Tecklenburg, M. and Lim, D. F. (1991) New Biol., 3, 799−811.
3. Blakely, G., Colloms, S., May, G., Burke, M., and Sherratt, D (1991) New Biol., 3, 789−798.

4. Sherratt, D., Blakely, G., Burke, M., Colloms, S., Leslie, N., McCulloch, R., May, G. and Roberts, J. (1993) In Helsop-Harrison, J. S. and Flavell, R. B. (eds), The Chromosome. BIOS Scientific Publishers, pp. 25–41.
5. Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992) Trends Genet. **8**, 432–439.
6. Summers, D. K. and Sherratt, D. J. (1984) Cell **36**, 1097–1103.
7. Colloms, S., Sykora, P., Szatmari, G. and Sherratt, D J. (1990) J. Bacteriol., **172**, 6973–6980.
8. Blakely, G., May, G , McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T. and Sherratt, D J. (1993) Cell **75**, 351–361.
9. Lovett, S. T. and Kolodner, R. D. (1991) J. Bacteriol., **173**, 353–364.
10. Missiakas, D., Georgopoulos, C. and Raina, S. (1994) EMBO J., **13**, 2013–2020.
11. Summers, D. K. and Sherratt, D. J. (1988) EMBO J., **7**, 851–858.
12. Stirling, C. J., Colloms, S. D., Collins, J. F., Szatmari, G. and Sherratt, D. J. (1989) EMBO J., **8**, 1623–1627.
13. Hoess, R. H., Wiezbicki, A and Abremski, K. (1986) Nucleic Acids Res., **14**, 2287–2301.
14. Bruckner, R. C. and Cox, M. M. (1986) J Biol. Chem , **261**, 11798–11807.
15. Summers, D. K. (1989) EMBO J., **8**, 309–315.
16. Truss, M., Chalepakis, G. and Beato, M. (1990) Proc Natl. Acad. Sci. USA., **87**, 7180–7184.
17. Wissmann, A. and Hillen, W. (1991) Methods Enzymol. **208**, 365–379
18. Siebenlist, U., Simpson, R. B. and Gilbert, W. (1980) Cell **20**, 269–281.
19. Pabo, C. O. and Sauer, R.T. (1992) Annu. Rev. Biochem., **61**, 1053–1095.
20. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. **65**, 499–560.
21. Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) Proc. Natl. Acad. Sci. USA., **73**, 804–808.
22. Panigrahi, G. B., Beatty, L. G. and Sadowski, P. D. (1992) Nucleic Acids Res., **20**, 5927–5935.
23. Harrison, S. C. and Aggarwal, A. K. (1990) Annu. Rev. Biochem., 59, 933–969.
24. Siebenlist, U. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA., **77**, 122–126.
25. Hillen, W., Schollmeier, K. and Gatz, C. (1984) J. Mol. Biol., **172**, 185–201.
26. Klemm, J. D., Rould, M A., Aurora, R., Herr, W. and Pabo, C. O. (1994) Cell **77**, 21–32.
27. Aggarwal, A. K., Rodgers, D. W , Drottar, M., Ptashne, M. and Harrison, S. C. (1988) Science **242**, 899–907.
28. Derbyshire, K. M. and Grindley, N. D. F (1992) EMBO J., **11**, 3449–3455
29. Pan, G. and Sadowski, P. D. (1993) J. Biol. Chem. **268**, 22546–22551.
30 Hendrikson, W. and Schleif, R. (1985) Proc. Natl. Acad. Sci. USA., **82**, 3129–3133.
31. Feng, J.-A., Johnson, R. C. and Dickerson, R. E (1994) Science **263**, 348–355.
32. Heuer, C. and Hillen, W. (1988) J. Mol. Biol., **202**, 407–415.
33. Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S., Pierson, L. S , Sternberg, N. and Leong, J. M. (1986) EMBO J. **5**, 433–440
34. Chen, J. -W., Lee, J. and Jayaram, M. (1992) Cell **69**, 647–658.
35. Abremski, K. E. and Hoess, R. H. (1992) Protein Engng **5**, 87–91.